U
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION (U)	1b. RESTRICTIVE MARKINGS NA
2a. SECURITY CLASSIFICATION AUTHORITY NA	3. DISTRIBUTION / AVAILABILITY OF REPORT Distribution Unlimited
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE NA	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) NA	5. MONITORING ORGANIZATION REPORT NUMBER(S) NA
6a. NAME OF PERFORMING ORGANIZATION California Institute of Technology 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Office of Naval Research
6c. ADDRESS (City, State, and ZIP Code) Division of Chemistry 147-75 California Institute of Technology Pasadena, CA 91125	7b. ADDRESS (City, State, and ZIP Code) 800 North Quincy Street Arlington, VA 22217-5000
8a. NAME OF FUNDING / SPONSORING ORGANIZATION DARPA/DSO 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-86-K-0775
8c. ADDRESS (City, State, and ZIP Code) 1400 Wilson Blvd. Arlington, VA 22209	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
Biopolymers: Proteins and Nucleic Acids

12. PERSONAL AUTHOR(S)
John H. Richards; John N. Abelson; Peter B. Dervan; Leroy H. Hood; Melvin I. Simon

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 9/15/86 TO 9/30/92	3/8/95	12

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Biopolymers/Macromolecular Interactions/Protein-DNA Binding/Peptide Synthesis/DNA Cleavage/NMR of Proteins and Nucleic Acids/tRNA Specificity

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The scientific work supported by N00014-86-K-0755 focused on the basis of information content in macromolecules and how this information directs various biological processes. Specifically we dealt with the activity of the two major classes of biological macromolecules, proteins and nucleic acids. We addressed such questions as: (i) how does the base sequence in a region of DNA influence its three-dimensional structure especially in ways that might facilitate the exquisitely specific recognition of a particular site in the DNA by a particular protein? (ii) how does the amino acid sequence of the protein that binds specifically to a particular site in DNA influence this binding and conversely how does this interaction depend on the sequence of the DNA? (iii) how can we use these insights to design proteins that will specifically recognize and cleave DNA? (iv) how does the base sequence in tRNA determine which amino acid will be specifically charged in a process that lies at the heart of the translation of the nucleic acid sequence of a gene into the unique amino acid sequence of the protein that the gene encodes?

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	(U)
22a. NAME OF RESPONSIBLE INDIVIDUAL Dr. Ira Skurnick & Dr. Robert W. Newburgh	22b. TELEPHONE (Include Area Code) 202/694-3147; 696-1212
	22c. OFFICE SYMBOL DARPA/DSO & ONR

DD Form 1473, JUN 86 *Previous editions are obsolete.* SECURITY CLASSIFICATION OF THIS PAGE

S/N 0102-LF-014-6603

FINAL REPORT

9/15/86-9/30/92

March 13, 1995

Biopolymers: Proteins and Nucleic Acids

California Institute of Technology
Pasadena, California 91125

Co-Investigators: John N. Abelson, Leroy E. Hood* and Melvin I. Simon, Division of Biology; Peter B. Dervan and John H. Richards, Division of Chemistry and Chemical Engineering

*Current address: Department of Molecular Biotechnology, University of Washington, Seattle, Washington

Contract: N00014-86-K-0755

Ph.D. Students Supported by Project
Michael Bruist
Thomas Chang
Michael Emerling
Kenneth S. Graham
Herman Hesen
Diane Hollenbaugh
Brent L. Iverson
David Long
David P. Mack
Tadashi (Jack) Mizoguchi
Jennifer Normanly
Martha G. Oakley
Todd Richmond
James P. Sluka
Claire Slutter
Yun Sun

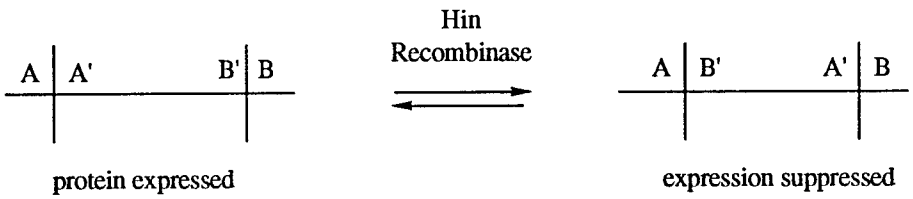
Postdoctoral Fellows Support by Project
Jamie E. Arenas
Wonhwa Cho
Anna C. Glasgow
Kelly T. Hughes
Heon Man Lim

Progress Report

The scientific work supported by N00014-86-K-0755 focused on the basis of information content in macromolecules and how this information directs various biological processes. Specifically we dealt with the activity of the two major classes of biological macromolecules, proteins and nucleic acids. We addressed such questions as: (i) how does the base sequence in a region of DNA influence its three-dimensional structure especially in ways that might facilitate the exquisitely specific recognition of a particular site in the DNA by a particular protein? (ii) how does the amino acid sequence of the protein that binds specifically to a particular site in DNA influence this binding and conversely how does this interaction depend on the sequence of the DNA? (iii) how can we use these insights to design proteins that will specifically recognize and cleave DNA? (iv) how does the base sequence in tRNA determine which amino acid will be specifically charged in a process that lies at the heart of the translation of the nucleic acid sequence of a gene into the unique amino acid sequence of the protein that the gene encodes?

The work has resulted in a number of papers and reviews (33 are submitted with this report) that cover these various issues and will be discussed with relation to the four questions just outlined.

The particular system on which we have focused the work of the first three questions plays a central role in a cell's response to its external environment, specifically response to concentration gradients of nutrients by *Salmonella typhimurium* as a manifestation of the general phenomenon of chemotaxis. The biological mechanism involves an inversion of a gene that encodes a protein essential for operation of the bacterial flagella. In one orientation the protein is expressed; in the other orientation expression is suppressed.



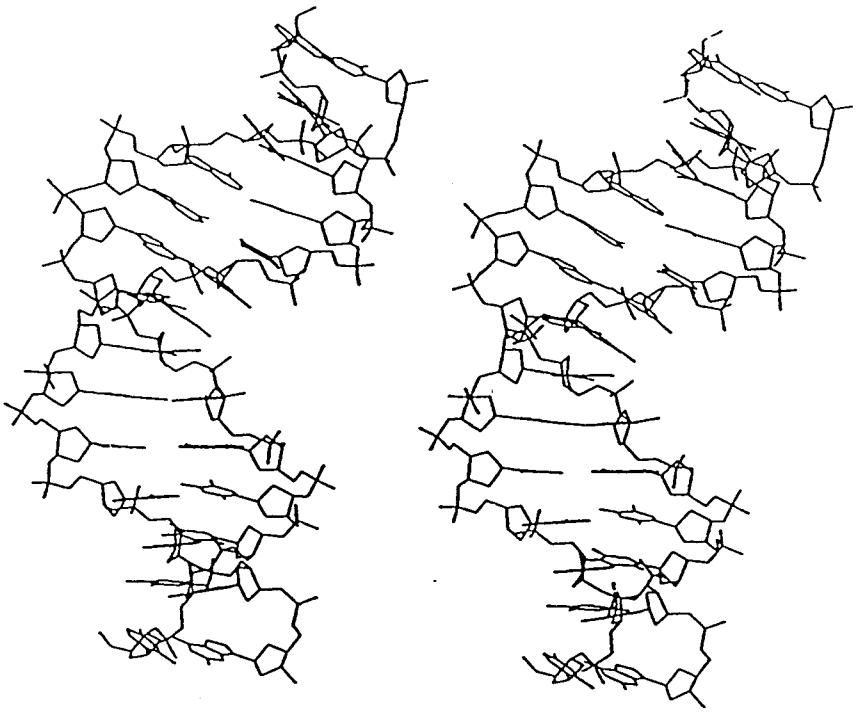
The system that accomplishes this rearrangement includes a number of proteins operating in concert. The most important of these proteins, Hin recombinase, specifically recognizes base sequences in the DNA labeled A, A', B, B' in the diagram and formally termed *hix* sites.

The fourth question focuses on the translation of a nucleic acid message into a protein. This process requires a complex biological machinery, one of whose essential elements is a family of transfer RNA species each of which accepts a particular amino acid which it then carries to the ribosome where that particular amino acid is inserted into the growing protein chain according to the sequence specified in the messenger RNA. The charging of a particular tRNA with its specific amino acid is accomplished by one of a class of enzymes that recognize certain structural features of a particular tRNA and reliably charge it with the appropriate amino acid. The act of charging each species of tRNA only with its cognate amino acid constitutes the central decoding step in translation of the information from the world of nucleic acids into the sequences of amino acids in a protein. The amino acid sequences of a protein in turn mandate the complex three-dimensional strucures and biological functions that characterize proteins.

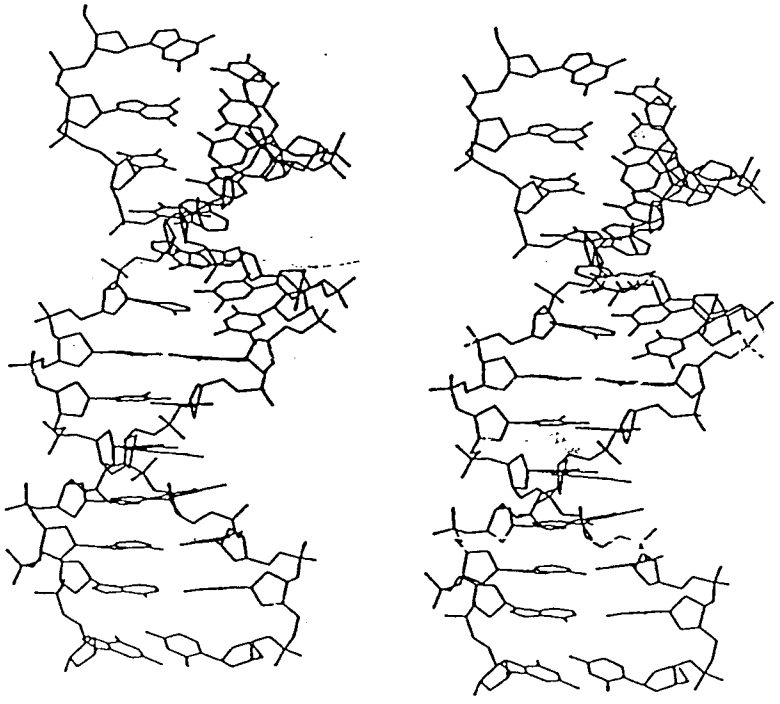
Together these biological phenomenon provide superb cases whose detailed analysis reveals the general features of the highly specific interactions between proteins and nucleic acids that provide the highly sophisticated recognition between the two most important classes of biological macromolecules.

(i) Sequence-dependent structure of DNA. Multi-dimensional magnetic resonance studies of two 14 base pair DNA molecules that contain the recognition sites for Hin recombinase (specifically R. *hix*L; d(GGTTTTTGATAAAG)•d(CTTTATCAAAAACC) and L. *hix*L; d(GGTTCTTGAAAACC)•d(GGTTTTCAAGAACC) show that these two sites have remarkably similar structures both of which differ appreciably from classical B-DNA. The refined DNA structures possess a significant bend (25–32°) in the middle of the helices. This bending causes a compression of the major groove at almost exactly the position where the recombinase binds. The DNA molecules were also found to have a deepened and narrowed minor groove near the continuous dA tracts, where the minor groove contacts occur between the N-terminal residues of the recombinase binding domain and the DNA molecules (Figure 1). Such pre-existing structural features of the free DNA molecules are likely to contribute to the specific interactions characteristic of Hin recombinase and the *hix* sites. Even slight changes in the base sequence of the 14 mers causes their structure to revert to that of classical B-DNA (1).

A



B



C

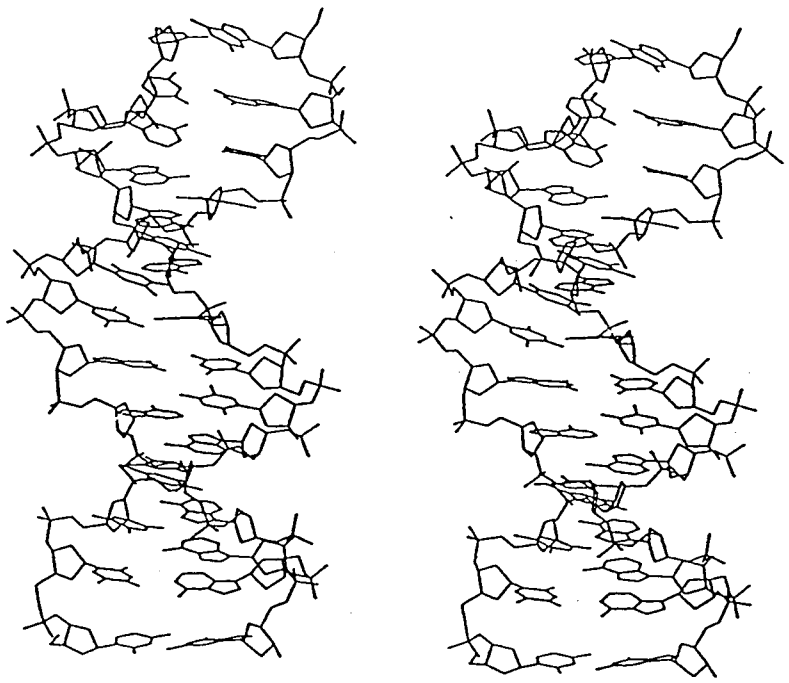


Figure 1.

Final structure of refined DNA R. *hix*L d(GGTTTTGATAAAG)•d(CTTTATCAAAAACC). Structures were generated by averaging coordinates in six converged structures. The resulting coordinates have been optimized for force field potentials. A. Side view: bending of the DNA molecule and the positive roll angle are shown. B. Rear view: at the junction between base-pairs T7•A22 and G8•C21, a positive slide has taken place so the interstrand Pur-Pur overlap can be maximized. C. Front view: narrowed major groove is shown.

Three-dimensional x-ray diffraction studies of the protein-nucleic complex have also been initiated (2).

(ii) How does the amino acids sequence of the protein that specifically binds to a particular site in DNA influence this interaction and conversely, how does this interaction depend on the structure of the DNA.

A study of phase variation in *Salmonella* analyzed the dependence of this biological phenomenon on the *in vivo* interaction between Hin recombinase and the *hix* recombination site. The bacteriophage P22-based challenge phase selection was used to characterize the binding of *Salmonella* Hin recombinase to the wild-type *hix*L or *hix*R recombination sites and to repress transcription from an upstream promoter in the challenge phase system. Hin-mediated repression results from Hin associating into multimers either prior to binding or during the cooperative binding process at the *hix* operator sites. The ability of Hin multimers to repress transcription is eliminated when the *hix* 13 bp half-sites are rotated to opposite sides of the DNA helix by inserting 4 bp between them. Insertion of 1 bp between half-sites reduces overall repression. Hin also binds one of the *hix*L half-sites to repress transcription, but only when high levels of Hin protein are present in the cell. Mutations have been identified in the *hix* sites that impair Hin binding. Five of the 26 bp in the *hix* sites are critical; sites with base-pair substitutions at these five positions show greatly reduced binding. Three additional base pairs make minor contributions to binding. These results are consistent with the results of binding studies between Hin and the *hix* sites *in vitro* (3).

Additional studies using nuclease and chemical protection further defined these interactions. The recombinase of the *Salmonella* inversion system, Hin, mediates site-specific recombination between two 26 bp inverted repeat sequences (*hix*L and *hix*R) which flank a 993 bp DNA segment. We investigated Hin recognition of, and association with, the *hix* recombination sites. Nuclease and chemical protection studies with linear and supercoiled DNA substrates demonstrate that Hin initially binds *hix*L and *hix*R independently of binding of the other protein components of the inversion system, Fis and HU. DNA-binding assays with mutant recombination sites and methylation interference experiments indicate that the critical bases for Hin recognition of its DNA-binding site are within an 8 bp sequence covering adjacent major and minor grooves of the DNA helix in each of the 12 bp half-sites of the *hix* recombination sites. The nature of the Hin-*hix* complexes in these binding studies and the results of gel filtration assays with purified Hin suggests that Hin binds the recombination sites as a dimer (4).

Use of DNA sequence variants that have pairs of single base substitutions reveal further details of the specificity of the Hin-*hix* recognition. The Hin recombinants of *Salmonella* catalyzes a site-specific recombination event which leads to flagellar phase variation. Starting with a fully symmetrical recombination site, *hix*C, a set of 40 recombination sites which vary by pairs of single base substitutions was constructed. This set was incorporated into the *Salmonella*-specific bacteriophage P22 based challenge phage selection and used to define the DNA sequence determinants for the binding of Hin to DNA *in vivo*. The critical sequence-specific contacts between a Hin monomer and a 13 bp *hix* half-site are at two T:A base pairs in the major groove of the DNA which are separated by one base pair, and two consecutive A:T contacts in the minor groove. The base substitutions in the major groove recognition protein which were defective in binding Hin still retained residual binding capability *in vivo*, while the base pair substitutions affecting the minor groove recognition region lost all *in vivo* binding. Using *in vitro* binding assays, Hin was found to bind to *hix* symmetrical sites with A:T base pairs or I:C base pairs in the minor groove recognition sequences, but not to G:C base pairs. In separate *in vitro* binding assays, Hin was equally defective in binding to either a G:C or a I:C contact in a major groove recognition sequence. Results from *in vitro* binding assays to *hix* sites in which 3-deazaadenione was substituted for adenine are consistent with Hin making a specific contact to either the N3 of adenine or O2 of thymine in the minor groove within the *hix* recombination site

on each symmetric half-site. These results taken with the results of previous studies on the DNA binding domain of Hin suggest a sequence-specific minor groove DNA binding motif (5).

As an aid in further analyses of this recognition, an artificial recombination site that is composed of two identical half sites that bind Hin recombinase was constructed and the effects of DNA mutations on recombination determined. An artificial recombination site *hix*C composed of two identical half-sites that bind the Hin recombinase served as a better operator *in vivo* than the wild type site *hix*L. *In vitro* binding assays such as gel retardation assay and methylation protection assay demonstrate that Hin binds to *hix*C as tightly as it binds to *hix*L, even when the sites are located in negatively supercoiled plasmids. However, *hix*C served as a poor recombination site when it was subjected to the standard inversion assay *in vitro*. *hix*C showed a 16-fold slower inversion rate than the wild type.

A series of biochemical assays designed to probe different stages of the Hin-mediated inversion reaction, demonstrated that Hin dimers bound to *hix*C have difficulty in forming paired *hix* site intermediates. $KMnO_4$ and S1 nuclease assays detected an anomalous structure of the center of *hix*C only when the site was in negatively supercoiled plasmids. Mutational analysis in the central region of *hix*C and assays of paired *hix* site formation with topoisomers of the *hix*C substrate plasmid suggest that Hin is not able to pair *hix*C sites because of the presence of the anomalous structure in the center of the site. The structure does not behave like a DNA "cruciform" since Hin dimers still bind efficiently to the site. It is thought to consist of a short denatured "bubble" encompassing 2 base pairs.

During the study of mutations in the center of *hix*C, we found that Hin is not able to cleave DNA if a guanine residue is one of the two central nucleotides close to the cleavage site. Furthermore, Hin acts in a concerted fashion and cannot cleave any DNA strand if one of the four strands in the inversion intermediate is not cleavable (6).

Interestingly supercoiling of the target DNA significantly influences the recombination process. A series of biochemical assays were developed and performed to monitor the molecular events that occur during the Hin-mediated DNA inversion reaction. These events can be divided into five different stages: (1) binding of proteins (Hin, Fis, and HU) to DNA; (2) pairing of Hin-binding sites; (3) invertasome formation; (4) dDNA strand cleavage; (5) strand rotation and religation. A series of topoisomers of the wild type DNA substrate plasmid [ranging from fully relaxed molecules to those with more than the physiological superhelical density (the physiological superhelical density of pKH336 from *E. coli* EH10B is –0.072 in this study)] was generated, and the role of negative supercoiling in each step of the inversion reaction was investigated. We found differences in the dependence of the formation of paired Hin-binding sites and of the invertasome formation on the superhelical density of the substrate plasmid. Pairing of Hin-binding sites occurs independently from invertasome formation, and a relatively low degree of negative supercoiling is enough to promote maximal pairing. However, efficient invertasome formation requires higher levels of negative supercoiling (7).

Minor groove contacts, as well as other interactions, have an important influence on the recognition of DNA by the binding domain of Hin recombinase. Incorporation of the DNA-cleaving moiety EDTA•Fe at discrete amino acid residues along a DNA-binding protein allows one to determine the positions of these residues relative to DNA bases, and hence the organization of the folded protein by mapping using high-resolution gel electrophoresis. A 52-residue protein, based on the sequence-specific DNA-binding domain of Hin recombinase (139-190), with EDTA at the amino terminus cleaves DNA at Hin recombination sites. The cleavage data for EDTA-Hin(139-190) reveal that the NH_2 terminus of Hin(139-190) is bound in the minor groove of DNA near the symmetry axis of Hin-binding sites . Six proteins varying in length from 49 to 60 residues and corresponding to the DNA-binding domain of Hin recombinase, were synthesized by solid-phase methods: Hin(142-190), Hin(141-190), Hin(140-

190), Hin(139-190), Hin(135-190), and Hin(131-190) were prepared with and without EDTA at the amino termini in order to test the relative importance of the residues Gly^{139}-Arg^{140}-Pro^{141}-Arg^{142}, located near the minor groove, for sequence-specific recognition at five imperfectly conserved 12-base-pair binding sites. Footprinting and affinity cleaving reveal that deletion of Gly^{139} results in a protein with affinity and specificity similar to those of Hin(139-190) but that deletion of Gly^{139}-Arg^{140} affords a protein with altered affinities and sequence specificities for the five binding sites. It appears that Arg^{140} in the DNA-binding domain of Hin is important for recognition of the 5'-AAA-3' sequence in the minor groove of DNA. Our results indicate modular DNA and protein interactions with two adjacent DNA sites (major and minor grooves, respectively) bound on the same face of the helix by two separate parts of the protein (8).

(iii) How can we use these insights to design proteins with attached functional groups that specifically bind and cleave DNA. Aspects of this subject have been reviewed in some breadth. High-resolution crystallographic views of protein-DNA complexes reveal the structural complexity of protein-DNA interactions. The combination of direct protein-DNA contacts mediated by multiple hydrogen bonds and sequence-dependent DNA conformational effects limits our ability to make detailed structural predictions, even if a new DNA-binding protein can be assigned to a structural class such as helix-turn-helix, double-barreled helix, zinc-binding finger, or scissor grip-leucine zipper. In the absence of high-resolution crystallographic and nuclear magnetic resonance (NMR) data, solution methods such as affinity cleaving can be used to characterize the topology of protein-DNA complexes and correlate sequence similarities with known structural classes (9).

The conversion of a sequence-specific DNA-binding protein into a sequence-specific DNA-cleaving protein by covalent attachment of the iron chelator, ethylenediamineetetraacetic acid (EDTA) to a specific amino acid residue creates a class of hybrid affinity-cleaving proteins that are available through chemical synthesis. Moreover, a structural domain consisting of naturally occurring amino acids that bind transition metals and oxidatively cleaves DNA extends this method to recombinant methods for protein synthesis.

The approach of affinity cleavage has been used to achieve this objective. In relation to the Hin recombinase problem this approach has involved synthetic 52-residue peptide based on the sequence-specific DNA-binding domain of Hin recombinase (139-190) that has been equipped with ethylenediaminetetraacetic acid (EDTA) at the amino terminus. In the presence of Fe(II), this synthetic EDTA-peptide leaves DNA at Hin recombination sites. The cleavage data reveal that the amino terminus of Hin(139-190) is bound in the minor groove of DNA near the symmetry axis of Hin recombination sites. This work demonstrates the construction of a hybrid peptide combining two functional domains: sequence-specific DNA binding and DNA cleavage (10).

A Hin DNA cleaving protein has also been used to reveal the orientation of the putative recognition helix in the DNA binding domains of Hin recombinase. On the basis of sequence similarity with other known DNA-binding proteins, the DNA-binding domain of Hin recombinase, residues 139-190, is thought to bind DNA by a helix-turn-helix motif. Two models can be considered that differ in the orientation of the recognition helix in the major groove of DNA. One is based on the orientation of the recognition helix found in the 434 repressor (1-69) and λ repressor-DNA cocrystals, and the other is based on the NMR studies of lac repressor headpiece. Cleavage by EDTA•Fe attached to a lysine side chain ($Ser^{183}\rightarrow Lys^{183}$) near the carboxyl terminus of Hin(139-184) reveals that the putative recognition helix is oriented toward the center of the inverted repeats in a manner similar to that seen in the 434 and λ repressor-DNA cocrystals (11).

Nickel has also been incorporated into a Hin cleaving protein to effect cleavage. A 55-residue protein containing the DNA binding domain of Hin recombinase, residues 139-190, with the tripeptide Gly-Gly-His (GGH) at the amino terminus was synthesized by stepwise solid-phase methods. GGH(Hin139-190) binds sequence specifically to DNA at four 134 base pair sites (termed *hix*L and *secondary*) and, in the presence of $Ni(OAc)_2$ and monoperoxyphthalic acid, reacts predominantly at a single deoxyribose position on one strand of each binding sites. We find that, upon treatment with n-butylamine, the DNA termini at the cleavage site are 3'- and 5'-phosphate, consistent with oxidative degradation of the deoxyribose backbone. The nickel-mediated oxidation can be activated with peracid, iodosylbenzene, or hydrogen peroxide. The sequence specificity of the reaction is not dependent on oxidant, but the rates of cleavage differ, decreasing in the order peracid > iodosylbenzene > hydrogen perocide. Optimal cleavage conditions for a 1 μM concentration of protein are 50 μM peracid, pH 8.0, and 1 equiv of $Ni(OAc)_2$. The preferential cleavage at a single base pair position on one strand of the minor groove indicates a nondiffusible oxidizing species. A change of absolute configuration in the GGH metal binding domain from L-His to D-His [Ni(II)•(GG-(-D-)H(Hin139-190)] affords cleavage at similar base pair locations but opposite with regard to strand specificity (12).

This approach has also been extended to the λδ-resolvase system as well as the *lac* operator. The DNA binding domain of λδ-resolvase, residues 141-183, is thought to bind DNA by a helix-turn-helix motif based on sequence similarities with other known DNA binding proteins. Incorporation of the DNA cleaving moiety, EDTA•Fe, at the NH_2 and COOH termini of λδ(141-183) allows the positions of these residues relative to the DNA bases at three resolvase binding sites, each consisting of inverted copies of an imperfectly conserved 9 bp sequence, to be mapped by high resolution gel electrophoresis. The cleavage data for EDTA-λδ(141-183) reveals that the amino terminus of the DNA binding domain of λδ-resolvase is bound proximal to the minor groove of DNA near the center of the resolvase binding sites. Cleavage by EDTA•Fe attached to a lysine side chain ($Asn^{183} \rightarrow Lys^{183}$) at the carboxyl terminus of λδ(141-183) reveals that the putative recognition helix is in the adjacent major groove on the same face of the helix, oriented toward the center of the inverted repeats (13).

Lac repressor (LacR) is a helix-turn-helix motif sequence-specific DNA binding protein. Based on proton NMR spectroscopic investigations, Kaptein and co-workers have proposed that the helix-turn-helix motif of LacR binds to DNA in an orientation opposite to that of the helix-turn-helix motifs of λ repressor, λ cro, 434 repressor, 434 cro, and CAP. In the present work, we have determined the orientation of the helix-turn-helix motif of LacR in the LacR-DNA complex by the affinity cleaving method. The DNA cleaving moiety EDTA•Fe was attached to the N-terminus of a 56-residue synthetic protein corresponding to the DNA binding domain of LacR. We have formed the complex between the modified protein and the left DNA half site for LacR. The locations of the resulting DNA cleavage positions relative to the left DNA half site provide strong support for the proposal of Kaptein and co-workers (14).

(iv) The relation between the base sequence of a given tRNA and the amino acid with which it becomes charged during protein synthesis is the central recognition event in translating the nucleic acid sequence in the mRNA into a specific amino acid sequence in the cognate protein. The structural basis for this fundamental recognition event has been studied by examining how changing certain bases in a tRNA can change the amino acid this mutated tRNA recognizes. In this context, changing 12 nucleotides transformed a leucine tRNA into a serine tRNA indicated that a limited set of residues determine tRNA identity (15). Later, a minimum number of 8 changes were found to be required to change a leucine amber suppressor codon into one recognizing serine. The appropriate changes were localized in the acceptor stem and in the D stem of the tRNA (16).

In pursuing these studies, amber suppressor genes encoding various amino acids had to be constructed. Specifically, amber suppressor genes corresponding to *E. coli* $tRNA^{Phe}$ and $tRNA^{Cys}$ were constructed for use in amino acid substitution studies as well as protein engineering. The genes for either $tRNA^{Phe}_{GAA}$ or $tRNA^{Cys}_{GCA}$ both with the anticodon 5' CTA 3' were assembled from four to six oligonucleotides, which were annealed and ligated into a vector. The suppressor genes are expressed constitutively from a synthetic promoter, derived from the promoter sequence of the *E. coli* lipoprotein gene. The $tRNA^{Phe}$ suppressor ($tRNA^{Phe}_{CUA}$) is 54–100% efficient *in vivo*, while the tRNA Cys suppressor ($tRNA^{Cys}_{CUA}$) is 17–50% efficient. To verify that the suppressors insert the predicted amino acids, both genes were used to suppress an amber mutation in a protein coding sequence. Amino-terminal sequence analysis of the resultant proteins revealed that $tRNA^{Phe}_{CUA}$ and $tRNA^{Cys}_{CUA}$ insert phenylalanine and cysteine, respectively. To demonstrate the potential of these suppressors, $tRNA^{Phe}_{CUA}$ and $tRNA^{Cys}_{CUA}$ have been used to effect amino acid substitutions at specific sites in the *E. coli lac* repressor (17).

Indeed a wider range of amber suppressor tRNA genes were constructed using synthetic oligonucleotides. In this regard, we have constructed 17 tRNA suppressor genes from *E. coli* representing 13 species of tRNA. We have measured the levels in *in vivo* suppression resulting from introducing each tRNA gene into *E. coli* via a plasmid vector. The suppressors function at varying efficiencies. Some synthetic suppressors fail to yield detectable levels of suppression, whereas others insert amino acids with greater than 70% efficiency. Results reported in the accompanying paper demonstrate that some of these suppressors insert the original cognate amino acid, whereas others do not. We have altered some of the synthetic tRNA genes in order to improve the suppressor efficiency of the resulting tRNAs. Both $tRNA^{His}_{CUA}$ and $tRNA^{Glu}_{CUA}$ were altered by single base changes, which generated –A–A– following the anticodon, resulting in a markedly improved efficiency of suppression. The $tRNA^{Pro}_{CUA}$ was inactive, but a hybrid suppressor tRNA consisting of the $tRNA^{Phe}_{CUA}$ anticodon stem and loop together with the remainder of the $tRNA^{Pro}$ proved highly efficient at suppressing nonsense codons. Protein chemistry results reported in the accompanying paper show that the altered $tRNA^{His}_{CUA}$ and the hybrid $tRNA^{Pro}_{CUA}$ insert only histidine and proline, respectively, whereas the altered $tRNA^{Glu}_{CUA}$ inserts principally glutamic acid but some glutamine. Also, a strain deficient in release factor 1 was employed to increase the efficiency of weak nonsense suppressors (18).

Bibliography and Papers Published (work referred to in text)

1. Sun, Y. NMR studies of protein-DNA interactions: Determinations of DNA structures recognized by Hin recombinase and studies of their roles in protein binding interactions. Ph.D. Thesis, California Institute of Technology (1992).

2. Feng, J.-A., Simon, M., Mack, D. P., Dervan, P. B., Johnson, R. C., and Dickerson, R. E. Crystallization and preliminary x-ray analysis of the DNA binding domain of Hin recombinase and its DNA binding site (1992).

3. Hughes K., Youderian, P., and Simon, M. K. Phase variation in *Salmonella*: Analysis of Hin recombinase and *hix* recombination site interaction *in vivo*. *Genes & Develop.* **2**, 937-48 (1988).

4. Glasgow, Z., Bruist, M. F., and Simon, M. I. DNA-binding properties of Hin recombinase. *J. Biol. Chem.* **264**, 10072-82 (1989).

5. Hughes, K. T., Gaines, P. C. W., Karlinsey, J. E., Vinayak, R., and Simon, M. I. Sequence-specific interaction of the *Salmonella* Hin recombinase in both major and minor grooves of DNA. *EMBO J.* **11**, 2695-2705 (1992).

6. Lim, H. M., Hughes, K. T., and Simon, M. I. The effects of symmetrical recombinase site *hixC* on Hin recombinase function. *J. Biol. Chem.* **267**, 11183-90 (1992).

7. Lim, H. M. and Simon, M. I. The role of negative supercoiling in Hin-mediated site-specific recombinase. *J. Biol. Chem.* **267**, 11176-82 (1992).

8. Sluka, J. P., Horvath, S. J., Glasgow, A. C., Simon, M. I., and Dervan, P. B. Importance of minor-groove contacts for recognition of DNA by the binding domain of Hin recombinase. *Biochemistry* **29**, 6551-61 (1990).

9. Dervan, P. B. Characterization of protein-DNA complexes by affinity cleaving. In *Methods in Enzymology* (Sauer, R., Ed.) Academic Press **208**, 497-515 (1991).

10. Sluka, J. P., Horvath, S. J., Bruist, M. F., Simon, M. I., and Dervan, P. B. Synthesis of a sequence-specific DNA-cleaving peptide. *Science* **238** 1129-32 (1987).

11. Mack, D. P., Sluka, J. P., Shin, J. A., Griffin, J. H., Simon, M. I., and Dervan, P. B. Orientation of the putative recognition helix in the DNA-binding domain of Hin recombinase complexed with the Hix site. *Biochemistry* **29**, 6561-7 (1990).

12. Mack, D. P. and Dervan, P. B. Sequence-specific oxidative cleavage of DNA by a designed metalloprotein, Ni(II)•GGH(Hin139–190) *Biochemistry* **31**, 9399-9405 (1992).

13 Graham, K. S. and Dervan, P. B. Structural motif of the DNA binding domain of λδ-resolvase characterized by affinity cleaving. *J. Biol. Chem.* **265**, 16534-40 (1990).

14. Shin, J. A., Ebright, R. H., and Dervan, P. B. Orientation of the Lac repressor DNA binding domain in complex with the left *lac* operator half site characterized by affinity cleaving. *Nucl. Acids. Res.* **19**, 5233-6 (1991).

15. Normanly, J., Ogden, R. C., Horvath, S. J., and Abelson, J. Changing the identity of a transfer RNA. *Nature* **321**, 213-9 (1986).

16, Normanly, J., Ollick, T., and Abelson, J. Eight base changes are sufficient to convert a leucine-inserting tRNA into a serine-inserting tRNA. *Proc. Natl. Acad. Sci. USA* **89**, 5680-4 (1992).

17. Normanly, J., Masson, J.-M., Kleina, L. G., Abelson, J., and Miller, J. H. Construction of two *Escherichia coli* amber suppressor genes: $tRNA^{Phe}_{CUA}$ **and** $tRNA^{Cys}_{CUA}$. *Proc. Natl. Acad. Sci. USA* **83**, 6548-52 (1986).

18. Kleina, L. G., Masson, J.-M., Normanly, J., Abelson, J., and Miller, J. H. Construction of *Escherichia coli* amber suppressor tRNA genes. II. Synthesis of additional tRNA genes and improvement of suppressor efficiency. *J. Mol. Biol.* **213**, 705-17 (1990).

Other Published Papers

19. Normanly, J., Kleina, L. G., Masson, J.-M., Abelson, J., and Miller, J. H. Construction of *Escherichia coli* amber suppressor tRNA genes. III. Determination of tRNA specificity. *J. Mol. Biol.* **213**, 719-26 (1990).

20. Arenas, J. E. and Abelson, J. The *Saccharomyces cerevisiae PRP21* gene product is an integral component of the prespliceosome. *Proc. Natl. Acad. Sci. USA* **90**, 6771-5 (1993).

21. Yamao, F., Inokuchi, H., Normanly, J., Abelson, J., and Ozeki, H. Mischarging mutants of Su^{+2} glutamine tRNA in *E. coli*. II. Amino acid specificities of the mutant tRNAs. *Jpn. J. Genet.* **63**, 251-8 (1988).

22. Masson, J.-M., Meuris, P., Grunstein, M., Abelson, J., and Miller, J. H. Expression of a set of synthetic suppressor tRNAPhe genes in *Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA* **84**, 6815-9 (1987).

23. Normanly, J. and Abelson, J. tRNA identity, *Ann. Rev. Biochem.* **58**, 1029-49 (1989).

24. Mack, D. P., Sluka, J. P., Shin, J. A., Griffin, J. H., Simon, M. I., and Dervan, P. B. Orientation of the putative recognition helix in the DNA-binding domain of Hin recombinase complexed with the Hix site. *Biochemistry* **29**, 6561-7 (1990).

25. Glasgow, A. C., Hughes, K. T., and Simon, M. I. Bacterial DNA inversion systems. In *Mobile DNA* (Berg, D. and Howe, M., Eds.) American Society for Microbiology, pp. 637-59 (1988).

26. Johnson, R. C., Ball, C. A., Pfeffer, D., and Simon, M. I. Isolation of the gene encoding the Hin recombinational enhancer binding protein. *Proc. Natl. Acad. Sci. USA* **75**, 3484-8 (1988).

27. Bruist, M. F., Glasgow, A. C., Johnson, R. C., and Simon, M. I. Fis binding to the recombinational enhancer of the Hin DNA inversion system. *Genes & Develop.* **1**, 762-62 (1987).

28. Johnson, R. C., Glasgow, A. C., and Simon, M. I. Spatial relationship of the Fis binding sites for Hin recombinational enhancer activity. *Nature* **329**, 462-5 (1987).

29. Johnson, R. C. and Simon, M. I. Enhancers of site-specific recombination in bacteria. *Trends in Genet.* **3**, 262-7 (1987).

30. Mack, D. P., Iverson, B. T., and Dervan, P. B. Design and chemical synthesis of a sequence-specific DNA-cleaving protein. *J. Am. Chem. Soc.* **110**, 7572-4 (1988).

31. Mack, D. P. and Dervan, P. B. Nickel-mediated sequence-specific oxidative cleavage of DNA by a designed metalloprotein. *J. Am. Chem. Soc.* **112**, 4604-6 (1990.

32. Oakley, M. G. and Dervan, P. B. Structural motif of the GCN4 DNA binding domain characterized by affinity cleaving. *Science* **248**, 847-50 (1990).

33. Sluka, J. P., Griffin, J. H., Mack, D. P., and Dervan, P. B. Reagents and methods for the solid-phase synthesis of protein-EDTA for use in affinity cleaving. *J. Am. Chem. Soc.* **112**, 6369-74 (1990).